

REVIEW ARTICLE

Open Access

Challenges in the introduction of next-generation sequencing (NGS) for diagnostics of myeloid malignancies into clinical routine use

Ulrike Bacher^{1,2}, Evgenii Shumilov³, Johanna Flach⁴, Naomi Porret¹, Raphael Joncourt¹, Gertrud Wiedemann¹, Martin Fiedler², Urban Novak⁵, Ursula Amstutz² and Thomas Pabst⁵

Abstract

Given the vast phenotypic and genetic heterogeneity of acute and chronic myeloid malignancies, hematologists have eagerly awaited the introduction of next-generation sequencing (NGS) into the routine diagnostic armamentarium to enable a more differentiated disease classification, risk stratification, and improved therapeutic decisions. At present, an increasing number of hematologic laboratories are in the process of integrating NGS procedures into the diagnostic algorithms of patients with acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), and myeloproliferative neoplasms (MPNs). Inevitably accompanying such developments, physicians and molecular biologists are facing unexpected challenges regarding the interpretation and implementation of molecular genetic results derived from NGS in myeloid malignancies. This article summarizes typical challenges that may arise in the context of NGS-based analyses at diagnosis and during follow-up of myeloid malignancies.

Biological-clinical challenges

Defining the clinical impact of novel NGS markers for different myeloid entities

Table 1 current diagnostics of myeloid malignancies, including acute myeloid leukemia (AML), myelodysplastic syndromes (MDS) and myeloproliferative neoplasms (MPN) has been rapidly evolving^{1,2}. Within the last 5–10 years, next-generation sequencing (NGS) has been introduced in most specialized hematologic laboratories with various myeloid NGS panels now being commercially available. These panels are based on targeted resequencing and usually analyze 25–50 genes. The genes tested

within these panels can be classified into several functional categories including the splicing machinery (e.g., *U2AF1*, *SF3B1*, *SRSF2*, *ZRSR2*), epigenetic modifiers (such as *TET2*, *DNMT3A*, *BCOR*, *ASXL1*, *IDH1*, *IDH2*), cohesins (*STAG2*, *RAD21*, and *SMC3*), transcription factors (*RUNX1*, *WT1*, *ETV6*), signaling molecules (*NF1*, *NRAS*, *CBL*, *PTPN11*, *JAK2*, *FLT3*), and chromatin modifiers (*EZH2*, *ASXL1*)^{3–6}.

With the help of myeloid gene panels more than one recurrent somatic mutation can be identified in most AML patients, and even within defined AML entities additional molecular genetic mutations are detectable in many cases⁷. In MDS, NGS allows the identification of molecular mutations in nearly 90% of the patients^{3,5,8–11}. As a result, molecular genetic markers find increasing entrance into current classification systems. In this context, the favorable subcategory of MDS with ring sideroblasts has been expanded to include cases with ≥5% of ring sideroblasts in the setting where a *SF3B1* mutation is

Correspondence: Ulrike Bacher (veraulrike.bacher@insel.ch) or Thomas Pabst (thomas.pabst@insel.ch)

¹Department of Hematology and Central Hematology Laboratory, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

²Center for Laboratory Medicine (ZLM)/University Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Full list of author information is available at the end of the article.

© The Author(s) 2018



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

present in the most recent revision of the WHO classification¹². Also for AML, new molecularly-defined entities have been suggested including chromatin–spliceosome, *TP53* aneuploidy, and provisionally, *IDH2*R172⁴.

The molecular genetic landscape undergoes exploration by NGS also in *BCR-ABL1*-negative MPNs^{13–16}. Besides the three classical mutations in *JAK2*, *CALR*, and *MPL* that are commonly referred to as MPN driver mutations, non-driver mutations in the genes known from MDS and AML are also detected in polycythemia vera (PV), essential thrombocythemia (ET), and primary myelofibrosis (PMF)^{15,17,18}. However, as compared to AML and MDS, the clinical significance of these novel markers has been more difficult to define. Considering the importance of cytomorphology and histopathology in addition to the

high frequency of the three driver mutations in the MPNs, which may also be investigated by traditional molecular techniques¹², the value of NGS at present lays in the refinement of risk stratification in critical or difficult cases and consequently treatment decisions^{19,20}. Along this line, the prognostic impact of certain co-occurring mutations has been associated with MPN disease progression as well as the development of secondary AML²¹. This is particularly relevant in PMF, where high molecular risk (HMR) markers include mutations in *ASXL1*, *SRSF2*, *EZH2*, *IDH1* and *IDH2*¹⁸. Similarly, in ET and PV, mutations with adverse prognostic value are *IDH2*, *U2AF1*, *EZH2*, *TP53*, *SH2B3*, and *SF3B1*¹⁴. The number of co-occurring mutations has also been suggested to be prognostically relevant with ≥ 2 HMR mutations predicting the worst

Table 1 Challenges accompanying the introduction of massive parallel sequencing in clinical routine diagnostics in hemato-oncology

Challenge	Background	Current and future approach
Discrimination of leukemia-related mutations from polymorphisms or passenger mutations	Driver mutations expected to occur at higher allele frequency in patient samples than passenger mutations; driver mutations more likely to have an impact on protein function than polymorphisms or passenger mutations	Optimization of cancer-specific databases including reporting of rare physiological gene variants Implementation of novel bioinformatic algorithms based on prediction of functional impact Quantitative and dynamic VAF monitoring (separately and together with other mutations) at follow-up
Discrimination of somatic leukemia-related mutations from CHIP	CHIP is presented in ~10% of individuals aged 70 to 80 and in up to 20% in the age group > 80 years	Quantitative and dynamic VAF monitoring (separately and together with other mutations) at follow-up Clarifying the significance of CHIP in the context of myeloid malignancies
Discrimination of leukemia-related somatic mutations from pathogenic germline alterations	Challenge to differentiate acquired somatic mutations from germline pathogenic variants at diagnosis	Mutation detection in germline control samples (e.g., skin fibroblasts, saliva) in mutations such as in <i>RUNX1</i> , <i>CEBPA</i> Thorough medical family history followed by molecular genetic tests in relatives if necessary High and stable VAF (e.g., 40–50%) at follow-up despite clinical response to treatment may be indicative for germline alteration
Discrimination of true genetic alterations from PCR, sequencing and post-sequencing artifacts	Many artefacts are known to arise during NGS library preparation, sequencing and data analysis	Error correction using molecular identifiers that individually label original input DNA molecules Refinement of error-correction computational methods in post-sequencing NGS data analysis Confirmation using Sanger sequencing
Limited sensitivity of NGS for minimal residual disease (MRD) assessment	Mutations detected at diagnosis may be re-identified at best to a VAF of 1–2%	Error-corrected sequencing using molecular identifiers Complementation of NGS by established MRD tools like real-time PCR and flow cytometry
High financial burden; demand on interdisciplinary approaches	Expensive technical and staff equipment, sophisticated data interpretation Complex translation of NGS results into therapeutic decisions	Development of continuously updated NGS interpretation sets and algorithms for well-established mutational profiles within distinct hematological malignancies Interdisciplinary leukemia boards

VAF variant allele frequency, CHIP clonal hematopoiesis of indeterminate significance, bp base pairs, G guanine, C cytosine, ITDs internal tandem duplication

Table 2A Overview on databases used for the characterization of genetic variants

Database	URL	Description
cBioPortal	www.cbioportal.org	Free database for cancer genomics
My Cancer Genome	www.mycancergenome.org	Free database for cancer genomics
COSMIC	cancer.sanger.ac.uk/cosmic	Catalogue Of Somatic Mutations In Cancer, free database for somatic mutations in cancer
dbSNP	www.ncbi.nlm.nih.gov/SNP	Free database for short genetic variations
ExAC Browser	exac.broadinstitute.org	Freely available exome sequencing data from the Exome Aggregation Consortium
ClinVar	www.ncbi.nlm.nih.gov/clinvar	Free database for information about genomic variations and their relationship to human health
gnomAD	gnomad.broadinstitute.org	Genome Aggregation Database, free database for genome sequencing data
ESP	evs.gs.washington.edu/EVS	Exome Sequencing Project, free database for exome sequencing data
LOVD	www.lovd.nl	Leiden Open Variation Database, freely available tool for gene-centered collection and display of DNA variations
HGMD Professional	www.hgmd.cf.ac.uk	Database for known gene mutations causing inherited diseases

prognosis and shortened leukemia-free survival²². Thus, NGS enables the identification of patients that are at higher risk for progression and transformation, whereas during follow-up, the occurrence of clonal molecular evolution or the detection of HMR marker can identify MPN patients that may become potential candidates for allogeneic hematopoietic stem cell transplantation (HSCT)²³.

Besides allowing a refinement of risk stratification and therapeutic decision making in patients with driver mutations, NGS offers great benefit in about 10% of patients with pathologic features of MPN that lack defining molecular drivers (i.e., *JAK2V617F*, *CALR*, and *MPL*)^{14,24}. Diagnostics in these so-called triple-negative (TN) cases can be challenging, so NGS may confirm the presence of hematological clonal disease and corroborate an initial cytomorphologic diagnosis. TN patients may carry driver mutations in non-canonical sites in *JAK2* and *MPL* or in alternative genes including epigenetic modifiers (*ASXL1*, *TET2*), the spliceosome (*SF3B1*, *SRSF2*), and regulators of cytokine signaling (*CBL*, *SH2B3*)¹⁴. However, despite growing implementation of myeloid NGS panels, a small proportion of TN MPN patients remains without any mutation detectable.

Together, the growing insights into molecular aspects of the pathogenesis of myeloid malignancies will eventually pave the way towards a more detailed clinical evaluation and optimized therapeutic decisions.

Discriminating leukemia-related mutations from genetic polymorphisms and passenger mutations

Following the expansion of large-scale cancer genome sequencing, databases such as Genome Aggregation Database (gnomAD), 1000-Genomes-Projekt and Exome Aggregation Consortium aim to depict population allele

frequencies in detail^{25–27} (Table 1). Other databases (Table 2A), e.g., International Agency for Research on Cancer, Catalog of Somatic Mutations in Cancer, Cancer Genome Atlas, and Human Gene Mutation Database focus on somatic and germline mutations to distinguish particular gene variants between “true” tumor associated mutations and genetic polymorphisms^{28–32}. Some of these databases are accessible via genome browser apps (e.g., Alamut). Based on such tools, gene variations described with frequencies of more than 1% in the population represent genetic polymorphisms.

In sight of the enormous number and variety of gene alterations detectable by NGS, the discrimination of leukemia-initiating mutations from incidental passenger mutations lacking any impact on leukemogenesis can be challenging. In fact, only a small fraction of somatic gene alterations seems to act as driver mutations that are involved in cancer initiation and progression^{33,34}. Different bioinformatic methods are utile to discriminate driver mutations from passenger variants, either by assessing the frequency of mutations or by predicting their functional impact^{35,36}. The first approach assumes a driver gene to be mutated in a significantly higher proportion of alleles compared to the expected background mutational rate and is calculated based on statistical scores (e.g., CancerMutationAnalysis, CHASM, DMI) taking into account, amongst others, gene size, nucleotide constitution and background non-synonymous mutation rates from cancer-specific databases (e.g., COSMIC)^{37–39}. The second approach is based on the hypothesis that mutations with “damaging” functional impact, predicted by tools like SIFT, MutationAssessor or MAPP, are more likely to represent driver mutations than those with little (if any) predicted impact^{35,36}. (Table 2B)

Table 2B Overview on algorithms used for the validation and interpretation of genetic variants

Algorithm	URL	Description
Align GVGD	agvgd.hci.utah.edu	Free web-based program for classification of missense variants
SIFT	sift.bii.a-star.edu.sg	Free web-based program for prediction of missense variant effect on protein function
MutationTaster	www.mutationtaster.org	Free web-based program for prediction of disease-causing potential of DNA variants
PolyPhen-2	genetics.bwh.harvard.edu	Free web-based program for prediction of missense variant effect on protein structure and function
Provean	provean.jcvi.org	Free web-based program for prediction of missense and indel variant effect on protein function
FATHMM	fathmm.biocompute.org.uk	Free web-based program for prediction of functional consequences of coding and non-coding variants

To further illustrate the difficulties that may arise in the discrimination of germline polymorphisms from “true” somatic mutations, we detected a c.167 T > C, p.(Leu56-Ser) alteration in the *RUNX1* gene in a patient with therapy-related AML. In the PB, the variant allele frequency (VAF) was 44% at diagnosis. At 9 months from diagnosis, the patient was in remission, however, the respective *RUNX1* alteration persisted at a VAF of 48% in the PB, whereas a co-incidental *SF3B1* mutation had decreased from 10.5 to 2.9% in the PB. At this later time point, we identified the aforementioned *RUNX1* alteration as a germline polymorphism occurring at a frequency of around 1.5% in the general population (gnomAD). This information had not been available at diagnosis of the AML. Thus, considering the steadily improvement of molecular genetic databases, population frequency data are expected to become more and more available even for very rare germline variants in the near future.

Nevertheless, the evidence underlying the respective databases is not without limitations and a satisfying definition of levels of evidence is lacking. Browsers that facilitate data aggregation and display such as Alamut require further optimization regarding these functions. Thus, the review of the reported genetic variants by experts with technical and clinical knowledge is highly relevant, and literature research remains mandatory, too. Interdisciplinary teams of expert physicians, genetic counselors, and technologists are required to improve the level of accuracy of genomic databases and the interpretation of distinct genetic variants that may occur in myeloid malignancies. In addition, further high quality clinical trials and functional studies are needed to improve our understanding of the relevance of less well-characterized genes and gene mutations.

Discrimination of somatic leukemia-related mutations from CHIP

The discrimination of somatic leukemia-associated mutations from clonal hematopoiesis of indeterminate potential (CHIP) is another challenging issue^{40,41}. Clonal hematopoiesis is typically benign in healthy individuals

with very small clones, while patients with clinically abnormal hematopoiesis, larger clones and more driver gene mutations appear to be at much greater risk⁴². CHIP is defined by evidence of a somatic mutation in a leukemia-associated driver gene amounting to an allele frequency of 2% or more in individuals that do not fulfill the WHO criteria for a hematologic malignancy⁴³. Kwok et al. identified somatic mutations in myeloid malignancy genes in 71% of patients with MDS, 62% of patients with ICUS (idiopathic cytopenia of unclear significance) and evidence of some dysplasia, and in 20% of ICUS patients without dysplasia. Variant allele fractions were comparable between patients with clonal ICUS and MDS⁴⁴.

In AML, *DNMT3A* mutations can persist in the post-therapeutic period despite continuous remission without affecting the relapse rate in the absence of co-incidental gene mutations^{45,46}. Aside from *DNMT3A*, the spectrum of CHIP may comprise a wide range of genes including *TET2*, *ASXL1*, *RUNX1*, *IDH1/2*, and others. Jongen-Lavrencic et al. recently demonstrated no adverse effect of persisting *DNMT3A*, *TET2* and *ASXL1* mutations in the absence of other genetic alterations in a cohort of 482 AML patients achieving first remission after two cycles of intensive induction treatment, thereby allocating them to CHIP rather than to a pre-relapse condition. Yet, any additional mutation present in the same patients with mutated *DNMT3A*, *TET2*, and *ASXL1* had a dramatic impact on the cumulative relapse incidence⁴⁷. Following this line of research, we recently reported a patient with *NPM1* mutation subtype switch at relapse of AML 8 years after successful intensive chemotherapy and consolidation with autologous stem cell transplantation (*NPM1*mut subtype D at first diagnosis; subtype A at relapse). Interestingly, at every assessment since diagnosis, a *DNMT3A* mutation was present at allele fractions between 37 and 51%⁴⁸. Furthermore, the *JAK2V617* mutation, which constitutes one of the most common mutations described in CHIP, occurs in about 0.1% of the general population without clinical signs of myeloproliferative disease^{49–51}. Therefore, caution should be used in order not to refer a patient automatically to a diagnosis of MPN in the case of

a *JAK2* mutation detected by traditional molecular techniques or NGS. This aspect emphasizes the need to clearly separate true MPNs from CHIP when a *JAK2* mutation has been detected, e.g., by considering the results of a bone marrow biopsy. Clearly, the role of clonal hematopoiesis requires careful investigation considering the long-term courses of patients. In addition, patients with unexplained thrombocytosis should not automatically be assigned to a diagnosis of MPN in case of detection of a mutation such as *DNMT3A* or *TET2* (that may occur in myeloid malignancies but may also reflect clonal hematopoiesis).

Naturally, while the differentiation between CHIP and leukemia-associated mutations is important, the finding of CHIP itself deserves attention as well. Allogeneic stem cell donation from individuals harboring CHIP, e.g., of *DNMT3A* may result in poor hematopoietic engraftment in the allograft recipients⁵². Individuals with CHIP have an increased risk for various myeloid malignancies as well as for cardiovascular disease due to accelerated arteriosclerosis, probably secondary to vascular inflammation driven by clonally derived monocytes/macrophages⁴³.

Discrimination of leukemia-related somatic mutations from rare pathogenic germline alterations

Considering the variety of mutations such as *TP53*, *RUNX1*, *GATA2*, *CEBPA*, or *ASXL1* that may arise in the context of AML or MDS, but may also occur as rare pathogenic germline variants, there is considerable risk for misinterpretation^{53–55}. The “IARC TP53 Database” contains more than several hundred *TP53* germline mutations, most of which are associated with the Li-Fraumeni cancer predisposition syndrome, do not show any hotspots regions and often occur within exons 2–11^{56,57}. Recently, germline *TP53* mutations were reported in 6 out of 107 patients with treatment-related AML highlighting their role in leukemogenesis after cancer treatment⁵⁸. Up to 11% of AML patients with biallelic *CEBPA* mutations, in fact, harbor a mutant *CEBPA* germline predisposition. In contrast to somatic *CEBPA* mutations, which cluster in the C-terminus of the gene, germline mutations mostly affect the N-terminus^{55,59,60}. *RUNX1* germline mutations are associated with familial platelet disorder with predisposition to AML, they mostly occur in a mono-allelic form and - in contrast to sporadic AML - represent the initiating molecular genetic event in leukemogenesis⁶¹. Thus, the search for a potential germline origin of the mutation may be justified when NGS detects an isolated *RUNX1* alteration with a mutation load around 50% in a given AML patient. Following the increasing implementation of NGS in routine cancer diagnostics, a higher frequency of germline mutations will be detectable in the future. Along this line, Drazer et al. identified pathogenic or likely

pathogenic variants in genes associated with hereditary hematopoietic malignancies in 21% of 360 patients with hematologic malignancies (*ANKRD26*, *CEBPA*, *DDX41*, *ETV6*, *GATA2*, *RUNX1* and *TP53*). In addition, they were able to show that especially mutations with VAF of more than 40% were more likely to be germline mutations as demonstrated by parallel germline tissue analyses⁶².

In the MPNs, complexity is added by the recent identification of a number of rare germline predisposition alleles^{49,63}. Affected genes include *TERT*, *SH2B3*, *TET2*, *ATM*, *CHEK2*, *PINT*, *FG11B*, *MECOM*, *TERT*, *JAK2* and *HBSL1-MYB*, respectively^{49,63}. The prognostic value of these heritable genetic polymorphisms is not fully understood and is thus not currently recommended be used for estimating the risk of developing an MPN. As NGS is unable to distinguish between germline and somatic origins unless paired germline/tumor DNA samples are evaluated the individual family's history needs to be carefully taken into account²⁴. Similarly, driver MPN mutations may also be present in rare forms of hereditary thrombocytosis or erythrocytosis^{64,65}, where cytoreductive therapy is not currently recommended²⁴.

In clinical practice, true germline control samples derived from non-hematopoietic, non-malignant tissues like skin biopsies are not always available. The examination of saliva provides an alternative (albeit with a lower safety level due to the possible contamination by hematopoietic cells). In many cases, and when population frequency databases do not provide definite information, only follow-up monitoring of a given mutation from the bone marrow or PB allows clarification. When a patient achieves complete remission, but the respective mutation persists at the previously high allele fraction, an interpretation as CHIP or germline mutation (especially in the case of a mutation load around 50%) seems more likely. In cases where different mutations occur at the primary diagnosis and all other mutations decrease under therapy, a single persisting mutation requires diligent re-evaluation. In such cases, the analysis of non-myeloid tissue is mandatory for exclusion of a hereditary origin of the respective alteration.

Detection of reciprocal rearrangements

Commercial myeloid NGS panels nowadays also enable the detection of reciprocal gene rearrangements. In the pre-NGS era, a limited number of reciprocal rearrangements underwent screening by PCR at diagnosis of AML, such as *PML-RARA*, *RUNX1-RUNX1T1*, and *CBFB-MYH11*. In addition to these frequent rearrangements (that are usually included in myeloid panels), current NGS panels allow the detection of a broad variety of rare reciprocal rearrangements in hemato-oncologic malignancies^{66,67}. In fact, over the past years more than 9000 new fusion genes - mostly interpreted as passenger events

by comparison with the data of the CancerGenome Atlas network - have emerged following the introduction of NGS⁶⁸. This reflects the increased sensitivity of NGS to discover subtle intra-chromosomal rearrangements whereas fluorescence in situ hybridization (FISH) may only detect exchanges of considerably larger chromosome segments. Accordingly, 75% of gene fusions revealed by NGS are related to intra-chromosomal rearrangements⁶⁸.

Stengel et al. investigated targeted RNA sequencing by NGS for the detection of reciprocal rearrangements in 58 AML cases with suspected novel fusions based on the detection of only one partner gene (*RUNX1*, *ETV6*, *PDGFRB*, *KMT2A*, *RARA*, *NPM1*, *MECOM*, *PDGFRA*, *BCOR*, *TET2*, *NUP98*) by chromosome banding analysis and FISH. The second partner gene could be identified in 59% of the patients by extended RNA sequencing⁶⁹. Another example are unusual variants of APL with pathogenic *X-RARA* or alternative *PML-RARA* fusions revealed by NGS, which traditional methods including RT-PCR failed to detect⁷⁰. In a patient with t-AML and a complex aberrant karyotype including a trisomy 8, we recently detected a cryptic reciprocal *RUNX1-CBFA2T3* rearrangement [corresponding to t(16;21)(q24;q22)] in a total of 25,000 NGS reads. Metaphase FISH using a *RUNX1* break apart probe confirmed the t(16;21). According to the literature, t(16;21) positive AML is frequently associated with previous radio-/chemotherapy, typically demonstrates additional chromosomal aberrations including trisomy 8, and patients may show eosinophilia⁷¹.

In cases with high transcript numbers, the correct diagnosis is easy to obtain, whereas low transcript numbers render it difficult to discriminate true genetic alterations from artifacts or passenger mutations according to the above-described principles of driver gene detection. Interphase FISH can confirm cryptic rearrangements below the detection level of chromosome banding analysis, if commercial probes for the involved genes/breakpoints are available. The combination with FISH allows excluding false positive results in the case of higher read counts. On the other hand, the addition of FISH increases the costs and the turn-around-time. The clinical impact of rare fusion transcripts detected only by FISH but not by chromosome banding deserves diligent investigation in the next years.

Technical challenges

Discrimination of true genetic alterations from PCR, sequencing, and post-sequencing artifacts

NGS procedures are associated with sequence errors due to artifacts that originate from library preparation, the sequencing process itself, or data analysis (e.g., read mapping, variant calling) resulting in incorrect calling of DNA bases or sequence variants^{72–74}. Such challenges

may result from the investigated genomic region itself with under- or overrepresentation of particular amplicons due to sequence-specific biases in target enrichment or sequencing efficiency (Table 1). Indeed, the depth of coverage across the target region can differ between library preparation methods, with amplicon-based protocols overall showing less uniform coverage compared to hybridization-capture-based target enrichment⁷⁵. In addition, distinct sets of nucleotides can be associated with poor sequencing performance (e.g., GGT or GGC patterns for Illumina technique or homopolymer regions for Ion Torrent)⁷⁶. Thus, NGS quality depends on the properties of the target sequence^{77–80}. However, sequence differences in NGS reads may also reflect PCR errors such as base misincorporations or rearrangements occurring in the process of massive and simultaneous amplification within multiple rounds of PCR^{81,82}. With the increasing use of molecular identifiers (i.e., molecular barcodes that individually label each DNA molecule from the original sample), artifacts arising both from PCR amplification or sequencing may be significantly reduced due to the use of consensus sequencing of single DNA molecules⁸³.

Incorrectly called sequence variants may also arise from digital processing during data analysis. This phase, also called post-sequencing NGS pipeline, represents a multistep and sequential process including quality control of raw sequence reads; aligning to a reference genome/assembly; post-alignment quality control and recalibration; identification of mutations (variant calling and genotyping); post-variant call/genotyping quality control; and finally data storage^{77,84,85}. Although all these processes include some error-correction procedures, further computational methods are necessary to reduce the sequential bias at this stage.

Besides further improvement of NGS-based assays and bioinformatic pipelines, additional Sanger sequencing may be helpful to discriminate errors caused by NGS from true genetic alterations. However, the limited sensitivity of Sanger sequencing of around 20% does not allow the investigation of sequence alterations detected by NGS at lower allele fractions occurring e.g., in MDS or AML subclones. Even with further improvement of NGS-based assays and bioinformatic pipelines, overdependence on bioinformatics pipelines should be avoided. For example, indel mutations can be incorrectly called. Visual inspection of the aligned sequencing data is advisable in cases where there is a suspicion of incorrect mutation calling by the used tool⁸⁶. Additionally, Sanger sequencing may be helpful to discriminate errors caused by NGS from true genetic alterations.

In this context, consideration needs to be given to the impact of the reference sequences used on read mapping, variant calling and annotation. In particular, variant annotation may differ depending on the transcript set

used^{87,88}. In addition, newer versions of the human reference genome (GRCh38) are incorporating additional genomic regions and are increasingly accounting for population variation through the inclusion of alternate haplotypes in order to reduce reference sequence bias in read mapping⁸⁹. Changes in the reference sequence may thus require re-validation of performance characteristics of NGS methods. This explains, for example, why GRCh37 is still in use, although GRCh38 has been published some time ago. Furthermore, future challenges include the incorporation of alternate haplotypes into variant calling algorithms⁸⁹.

Other technical aspects

The currently used NGS technologies in clinical diagnostics usually are classifiable as short read sequencing (read length ~100–~500 bp). Short read sequencing is inherently prone to miss structural variants such as longer insertions and deletions. This holds true especially for capture based target enrichment as opposed to targeted library amplification methods as there may be insufficient positioning data to align the obtained reads to the reference correctly^{90–92}. Paired end sequencing can mitigate this issue by allowing sequencing of the same read from both ends providing additional positioning information compared with single end sequencing^{90,93}. This method of sequencing comes at a cost of more complex library preparation and data analysis, higher sequencing times and data volumes^{90–92}. With the ongoing maturation and future establishment of long read sequencing technologies in clinical diagnostics further tools will be available to complement the current short read sequencing applications. Long read sequencing can generate read lengths exceeding several kilobases. As such, it is much better suited to analyze structural variants and highly repetitive sequences. Additionally, it will much more readily allow direct phasing of concurrent genetic variants (cis/trans). Currently, however, the high costs of such sequencers still largely limits them to research settings^{90–92}.

Minimal residual disease

The multitude of molecular markers in AML detectable by standard myeloid NGS panels (e.g., *RUNX1*, *EZH2*, or spliceosome mutations), offers the possibility of performing molecular minimal residual disease (MRD) monitoring in virtually every patient. Most AML patients show more than one molecular marker⁴. However, MRD detection is limited by the background noise of NGS^{94,95}. Already known mutations are re-identifiable at best at a variant allele fraction as low as 1%. Thus, the sensitivity of MRD detection by NGS is limited compared to quantitative real-time PCR (qPCR) offering sensitivities of 10^{-4} to 10^{-6} , or the sometimes even more sensitive digital droplet PCR (ddPCR). Currently, qPCR or ddPCR are

primarily used for monitoring of frequently occurring genetic alterations, e.g., hotspot mutations such as *NPM1* subtype A, or recurrent reciprocal rearrangements such as *RUNX1-RUNX1T1*/t(8;21). Therefore, increasing the number of genes within a molecular MRD marker panel is of paramount importance. An additional challenge is the frequently polyclonal character of AML, as different clones may respond differently to therapy or may re-emerge separately in the post-therapeutic or post-transplant period. Error-corrected or barcoded sequencing using molecular identifiers increasing the sensitivity of NGS to around 10^{-5} may facilitate future MRD strategies⁹⁶. At present, a combination, e.g., of qPCR/ddPCR and NGS in the case of suitable marker combinations should be explored⁴⁷. Given the variety of the molecular markers and their combinations, evaluation of molecular MRD results often remains individualized and requires a close interaction with clinicians. Similarly, molecular MRD results should be correlated with peripheral blood values and the results of flow cytometry, cytomorphology, and histopathology, eventually also with cytogenetics/FISH.

When AML patients develop an overt or molecular relapse, clinicians should be alert to initiate comprehensive mutation screening again. This allows for the detection of therapeutic targets such as *IDH1*, *IDH2*, or *FLT3*-ITD/TKD that may have newly developed due to clonal evolution.

Economic and organizational aspects

Finally, the additional financial burden by the use of NGS deserves some consideration. Expenses concerning technical and staff equipment, including a variety of professionals such as molecular and computational biologists, genetic counsellors, and specialized clinicians, are considerably higher than anticipated some years ago. Even though the costs of a single NGS assay will continue to decrease, the number of genes that are required for comprehensive diagnostic results and treatment decisions is continuously growing. Therefore, the time allocation for the post-sequencing steps such as filtering of variants, comparison of data with genomic databases, and interpretation of gene variants as well as writing of laboratory reports has become a challenge. In addition, high in-depth knowledge and experience is required to meet all the demands of NGS procedures used in hematology with its extensive gene/hot-spot panels for the different hematologic subtypes.

At present, therapeutic concepts focus on allocating AML patients at the earliest possible time point to targeted therapies such as *IDH1*, *IDH2*, or *FLT3* inhibition. Considering the long turnaround time of NGS, this is a relevant challenge for molecular laboratories. Thus, in many instances, molecular laboratories perform

traditional assays (e.g., fragment analysis for *FLT3*-ITD or -TKD) in parallel to NGS, even though NGS provides the same results a few days later. This further increases the efforts and the costs for molecular analyses in leukemia patients. “Best-practice algorithms” that take labor and equipment resources into account and that aim to avoid duplicate analyses are needed.

Interdisciplinary approaches

For a correct interpretation of the variety of NGS results in the context of hematological diagnostics, a constant bidirectional interaction between laboratory scientists, technicians and clinicians is essential. This holds true for the selection of the appropriate NGS panels at diagnosis, the selection of markers for follow-up analyses in accordance to the suspected hematological malignancy, but also to the interpretation of the genetic alterations detected by modern sequencing approaches. Specifically, the results of NGS should be interpreted in the context of other laboratory findings regarding cytomorphology, histopathology, immunophenotyping, traditional molecular genetics, cytogenetics, and clinical diagnostic and data outcome. NGS-based MRD approaches in principle are more comprehensive in light of the variety of mutations detectable at AML diagnosis today, but limited resources require a diligent selection of the optimal time points and the markers for NGS follow-up and a comparison with traditional and currently more sensitive approaches (e.g., real-time PCR). Thus, informal and formal interdisciplinary communication (i.e., in the context of leukemia boards, molecular tumor boards) has shown to be particularly useful.

Conclusions

NGS has opened new horizons for individualized diagnostics and therapy of myeloid malignancies. While new technological advances may improve the sensitivity and accuracy of NGS-based analyses, its results also deserve cautious interpretation considering the clinical context. In particular, the clinical value of NGS needs to be defined for the different myeloid entities. For AML, the additional value of comprehensive analysis by NGS has to be compared to the previously limited molecular panels including some driver mutations only, such as *NPM1* or *FLT3*-ITD. The identification of additional prognostically relevant mutations such as the adverse *RUNX1* mutations or the identification of novel therapeutic targets such as *IDH1* or *IDH2* justify the additional demand that arises from comprehensive NGS analysis⁹⁷. For MDS patients, molecular analysis allows a more accurate prognostication as compared to cytogenetics alone⁵. Nevertheless, the WHO classification of myelodysplastic syndromes so far only respects the *SFB3B1* mutation¹² due to the overlap of MDS-related mutations with CHIP. In the MPNs, NGS

can confirm a clonal disorder in triple-negative cases and improves the characterization of the risk profile, e.g., in patients with PMF. Presence of two or more high molecular risk markers such as mutations in *ASXL1*, *EZH2*, or *IDH1/IDH2* predicts a highly adverse prognosis and rapid leukemic transformation²² and thus may identify potential candidates for allogeneic HSCT²³.

Apparent AML or MDS-associated mutations may reflect germline variants or clonal hematopoiesis. In critical cases, exclusion of a germline origin by analysis of buccal cells, or, ideally, a skin biopsy, is mandatory. Persistence of mutations such as *ASXL1*, *DNMT3A*, or *TET2* despite clearance of other mutations in AML patients under therapy should raise alertness for the possibility of clonal hematopoiesis⁴⁷. Molecular MRD diagnostics is likely to undergo an expansion in the future as in virtually all AML patients there are markers suitable for NGS analysis. The possibility of discordant dynamics of simultaneous mutations requires comprehensive MRD monitoring combining different techniques (mainly quantitative PCR and NGS). Comprehensive MRD panels targeting all initial markers allows for the detection of dissociated responses of polyclonal disease to therapy with some clones responding while others are refractory. Finally, clinicians should be alert to initiate comprehensive mutation screening in patients with molecular or overt relapse for detecting therapeutic targets such as *IDH1*, *IDH2*, or *FLT3*-ITD/TKD that may newly develop consequently to clonal evolution.

Taken together, hemato-oncologists and pathologists should remain in close interaction with laboratory specialists. An interdisciplinary approach is essential to avoid misinterpretation of results. Both clinicians and molecular biologists require in-depth training for the critical interpretation of NGS results, and their interdisciplinary communication is more essential than ever. Such combined efforts will contribute to an optimized use of these novel diagnostic methods for the benefit of patients with myeloid malignancies.

Author details

¹Department of Hematology and Central Hematology Laboratory, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. ²Center for Laboratory Medicine (ZLM)/University Institute of Clinical Chemistry, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland. ³Department of Hematology and Medical Oncology, University Medicine Göttingen (UMG), Göttingen, Germany. ⁴Department of Hematology and Oncology, Medical Faculty Mannheim of the Heidelberg University, Mannheim, Germany. ⁵Department of Medical Oncology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 June 2018 Revised: 17 September 2018 Accepted: 15 October 2018

Published online: 12 November 2018

References

- Shumilov, E. et al. Current status and trends in the diagnostics of AML and MDS. Review article. *Blood Rev.* (2018). <https://doi.org/10.1016/j.blre.04.008>.
- Barbui, T. et al. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. *Blood Cancer J.* **8** (2018). <https://doi.org/10.1038/s41408-018-0054-y>.
- Cazzola, M., Della Porta, M. G. & Malcovati, L. The genetic basis of myelodysplasia and its clinical relevance. *Blood* **122**, 4021–4034 (2013).
- Papaemmanuil, E. et al. Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
- Papaemmanuil, E. et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627 (2013).
- Duncavage, E. J. & Tandon, B. The utility of next-generation sequencing in diagnosis and monitoring of acute myeloid leukemia and myelodysplastic syndromes. *Int. J. Lab. Hematol.* **37**(Suppl 1), 115–121 (2015).
- Cancer Genome Atlas Research. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
- Yoshida, K. et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
- Bejar, R. et al. Clinical effect of point mutations in myelodysplastic syndromes. *N. Engl. J. Med.* **364**, 2496–2506 (2011).
- Haferlach, T. et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**, 241–247 (2014).
- Bacher, U., Kohlmann, A. & Haferlach, T. Mutational profiling in patients with MDS: ready for every-day use in the clinic? *Best. Pract. Res. Clin. Haematol.* **28**, 32–42 (2015).
- Arber, D. A. et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).
- Tefferi, A. et al. Targeted deep sequencing in primary myelofibrosis. *Blood Adv.* **1**, 105–111 (2016).
- Tefferi, A. et al. Targeted deep sequencing in polycythemia vera and essential thrombocythemia. *Blood Adv.* **1**, 21–30 (2016).
- Lundberg, P. et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* **123**, 2220–2228 (2014).
- Ortmann, C. A. et al. Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med.* **372**, 601–612 (2015).
- Tefferi, A. et al. Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood* **124**, 2507–2513 (2014).
- Vannucchi, A. M. et al. Mutations and prognosis in primary myelofibrosis. *Leukemia* **27**, 1861–1869 (2013).
- Tefferi, A. & Barbui, T. Polycythemia vera and essential thrombocythemia: 2017 update on diagnosis, risk-stratification, and management. *Am. J. Hematol.* **92**, 94–108 (2017).
- Tefferi, A. Primary myelofibrosis: 2017 update on diagnosis, risk-stratification, and management. *Am. J. Hematol.* **91**, 1262–1271 (2016).
- Vainchenker, W. & Kralovics, R. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood* **129**, 667–679 (2017).
- Guglielmelli, P. et al. The number of prognostically detrimental mutations and prognosis in primary myelofibrosis: an international study of 797 patients. *Leukemia* **28**, 1804–1810 (2014).
- Alduaij, W. et al. Clinical utility of next-generation sequencing in the management of myeloproliferative neoplasms: a single-center experience. *HemaSphere* **2**, e44 (2018).
- Rumi, E. & Cazzola, M. Advances in understanding the pathogenesis of familial myeloproliferative neoplasms. *Br. J. Haematol.* **178**, 689–698 (2017).
- Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
- Genomes Project. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Song, W. et al. Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. *Genet. Med.* **18**, 850–854 (2016).
- Cancer Genome Atlas Research. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
- Futreal, P. A. et al. A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
- International Cancer Genome. International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
- Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
- Stenson, P. D. et al. The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
- Pon, J. R. & Marra, M. A. Driver and passenger mutations in cancer. *Annu. Rev. Pathol.* **10**, 25–50 (2015).
- Zhang, J. et al. Identifying driver mutations from sequencing data of heterogeneous tumors in the era of personalized genome sequencing. *Brief. Bioinform.* **15**, 244–255 (2014).
- Carter, H. et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res.* **69**, 6660–6667 (2009).
- Tan, H., Bao, J. & Zhou, X. A novel missense-mutation-related feature extraction scheme for “driver” mutation identification. *Bioinformatics* **28**, 2948–2955 (2012).
- Wong, W. C. et al. CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. *Bioinformatics* **27**, 2147–2148 (2011).
- Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
- Xie, M. et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
- Bejar, R. CHIP, ICUS, CCUS and other four-letter words. *Leukemia* **31**, 1869–1871 (2017).
- Steensma, D. P. Clinical Implications of Clonal Hematopoiesis. *Mayo Clin. Proc.* **93**, 1122–1130 (2018).
- Kwok, B. et al. MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance. *Blood* **126**, 2355–2361 (2015).
- Ploen, G. G. et al. Persistence of DNMT3A mutations at long-term remission in adult patients with AML. *Br. J. Haematol.* **167**, 478–486 (2014).
- Bhatnagar, B. et al. Persistence of DNMT3A R882 mutations during remission does not adversely affect outcomes of patients with acute myeloid leukaemia. *Br. J. Haematol.* **175**, 226–236 (2016).
- Jongen-Lavrencic, M. et al. Molecular minimal residual disease in acute myeloid leukemia. *N. Engl. J. Med.* **378**, 1189–1199 (2018).
- Bacher, U. et al. Pitfalls in the molecular follow-up of NPM1 mutant acute myeloid leukemia. *Haematologica* (2018). <https://doi.org/10.3324/haematol.192104>.
- Hinds, D. A. et al. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood* **128**, 1121–1128 (2016).
- Nielsen, C., Birgens, H. S., Nordestgaard, B. G. & Bojesen, S. E. Diagnostic value of JAK2 V617F somatic mutation for myeloproliferative cancer in 49 488 individuals from the general population. *Br. J. Haematol.* **160**, 70–79 (2013).
- Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
- Gibson, C. J. et al. Donor-engrafted CHIP is common among stem cell transplant recipients with unexplained cytopenias. *Blood* **130**, 91–94 (2017).
- Churpek, J. E. et al. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. *Blood* **126**, 2484–2490 (2015).
- Feurstein, S., Drazer, M. W. & Godley, L. A. Genetic predisposition to leukemia and other hematologic malignancies. *Semin. Oncol.* **43**, 598–608 (2016).
- Pabst, T., Eyholzer, M., Haefliger, S., Schardt, J. & Mueller, B. U. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. *J. Clin. Oncol.* **26**, 5088–5093 (2008).

56. Kappel, S. et al. TP53 germline mutation may affect response to anticancer treatments: analysis of an intensively treated Li-Fraumeni family. *Breast Cancer Res. Treat.* **151**, 671–678 (2015).
57. Ruijs, M. W. et al. TP53 germline mutation testing in 180 families suspected of Li-Fraumeni syndrome: mutation detection rate and relative frequency of cancers in different familial phenotypes. *J. Med. Genet.* **47**, 421–428 (2010).
58. Zebisch, A. et al. Acute myeloid leukemia with TP53 germ line mutations. *Blood* **128**, 2270–2272 (2016).
59. Taskesen, E. et al. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. *Blood* **117**, 2469–2475 (2011).
60. Tawana, K. et al. Disease evolution and outcomes in familial AML with germline CEBPA mutations. *Blood* **126**, 1214–1223 (2015).
61. Bellissimo, D. C. & Speck, N. A. RUNX1 Mutations in Inherited and Sporadic Leukemia. *Front. Cell. Dev. Biol.* **5**, 111 (2017).
62. Drazer, M. W. et al. Prognostic tumor sequencing panels frequently identify germ line variants associated with hereditary hematopoietic malignancies. *Blood Adv.* **2**, 146–150 (2018).
63. Tapper, W. et al. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat. Commun.* **6**, 6691 (2015).
64. Kapralova, K. et al. Cooperation of germ line JAK2 mutations E846D and R1063H in hereditary erythrocytosis with megakaryocytic atypia. *Blood* **128**, 1418–1423 (2016).
65. Lanikova, L. et al. Coexistence of gain-of-function JAK2 germ line mutations with JAK2V617F in polycythemia vera. *Blood* **128**, 2266–2270 (2016).
66. Talkowski, M. E. et al. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am. J. Hum. Genet.* **88**, 469–481 (2011).
67. Wang, E. et al. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin. Cancer Biol.* **30**, 4–12 (2015).
68. Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer* **15**, 371–381 (2015).
69. Stengel, A. et al. Detection of recurrent and of novel fusion transcripts in myeloid malignancies by targeted RNA sequencing. *Leukemia* **32**, 1229–1238 (2018).
70. Welch, J. S. et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. *JAMA* **305**, 1577–1584 (2011).
71. De Braekeleer, E. et al. RUNX1-MTG16 fusion gene in acute myeloblastic leukemia with t(16;21)(q24; q22): case report and review of the literature. *Cancer Genet. Cytogenet.* **185**, 47–50 (2008).
72. Glenn, T. C. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* **11**, 759–769 (2011).
73. Sleep, J. A., Schreiber, A. W. & Baumann, U. Sequencing error correction without a reference genome. *BMC Bioinforma.* **14**, 367 (2013).
74. Shiroguchi, K., Jia, T. Z., Sims, P. A. & Xie, X. S. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proc. Natl Acad. Sci. USA* **109**, 1347–1352 (2012).
75. Samorodnitsky, E. et al. Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. *Hum. Mutat.* **36**, 903–914 (2015).
76. Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P. & Tyson, G. W. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS. Comput. Biol.* **9**, e1003031 (2013).
77. Sims, D., Sudbery, I., Illott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–132 (2014).
78. Abnizova, I. et al. Statistical comparison of methods to estimate the error probability in short-read Illumina sequencing. *J. Bioinform. Comput. Biol.* **8**, 579–591 (2010).
79. Shin, S. & Park, J. Characterization of sequence-specific errors in various next-generation sequencing systems. *Mol. Biosyst.* **12**, 914–922 (2016).
80. Nakamura, K. et al. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* **39**, e90 (2011).
81. Clarke, L. A., Rebelo, C. S., Goncalves, J., Boavida, M. G. & Jordan, P. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. *Mol. Pathol.* **54**, 351–353 (2001).
82. Fox, E. J., Reid-Bayliss, K. S., Emond, M. J. & Loeb, L. A. Accuracy of next generation sequencing platforms. *Next Gener. Seq. Appl.* **1** (2014).
83. Salk, J. J., Schmitt, M. W. & Loeb, L. A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat. Rev. Genet.* **19**, 269–285 (2018).
84. Dolled-Filhart, M. P., Lee, M. Jr., Ou-Yang, C. W., Haraksingh, R. R. & Lin, J. C. Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *ScientificWorldJournal* **2013**, 730210 (2013).
85. Wolfinger, M. T., Fallmann, J., Eggenhofer, F. & Amman, F. ViennaNGS: A toolbox for building efficient next-generation sequencing analysis pipelines. *F1000Res.* **4**, 50 (2015).
86. Roy, S. et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines. *J. Mol. Diagn.* **20**, 4–27 (2018).
87. Frankish, A. et al. Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. *BMC Genom.* **16**, S2–S2 (2015).
88. McCarthy, D. J. et al. Choice of transcripts and software has a large effect on variant annotation. *Genome Med.* **6**, 26–26 (2014).
89. Schneider, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).
90. Baker, S. C. Next-generation sequencing challenges. *GEN.* <https://www.genengnews.com/issue/toc/286>. (2017). accessed on May 20, 2018.
91. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: 10 years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333 (2016).
92. Mardis, E. R. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* **12**, 213 (2017).
93. Paired-end vs. single-read sequencing. *Illumina.* <https://emea.illumina.com/science/technology/next-generation-sequencing/paired-end-vs-single-read-sequencing.html>? (2018). accessed on May 15, 2018.
94. Cruz, N. M., Mencia-Trinchant, N., Hassane, D. C. & Guzman, M. L. Minimal residual disease in acute myelogenous leukemia. *Int. J. Lab. Hematol.* **39**(Suppl 1), 53–60 (2017).
95. Ommen, H. B. Monitoring minimal residual disease in acute myeloid leukaemia: a review of the current evolving strategies. *Ther. Adv. Hematol.* **7**, 3–16 (2016).
96. Young, A. L. et al. Quantifying ultra-rare pre-leukemic clones via targeted error-corrected sequencing. *Leukemia* **29**, 1608–1611 (2015).
97. Dohner, H. et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* **129**, 424–447 (2017).